

資料4

科学技術・学術審議会学術分科会
学術情報委員会（第7回）
平成30年2月1日（水）

オープンサイエンスを支える研究基盤

－ 国際動向とNIIが開発する研究データ基盤の開発状況 －

国立情報学研究所

山地一禎

2018年2月1日 文部科学省学術情報委員会

話の流れ

海外における研究データ基盤の整備状況

①



イギリス



ヨーロッパ



オーストラリア



アメリカ



②

- 世界のデファクトは？
- 日本はどうあるべきか？

③

NIIが開発する研究データ基盤の開発状況



GakuNin RDM



WEKO3




CiNii Research

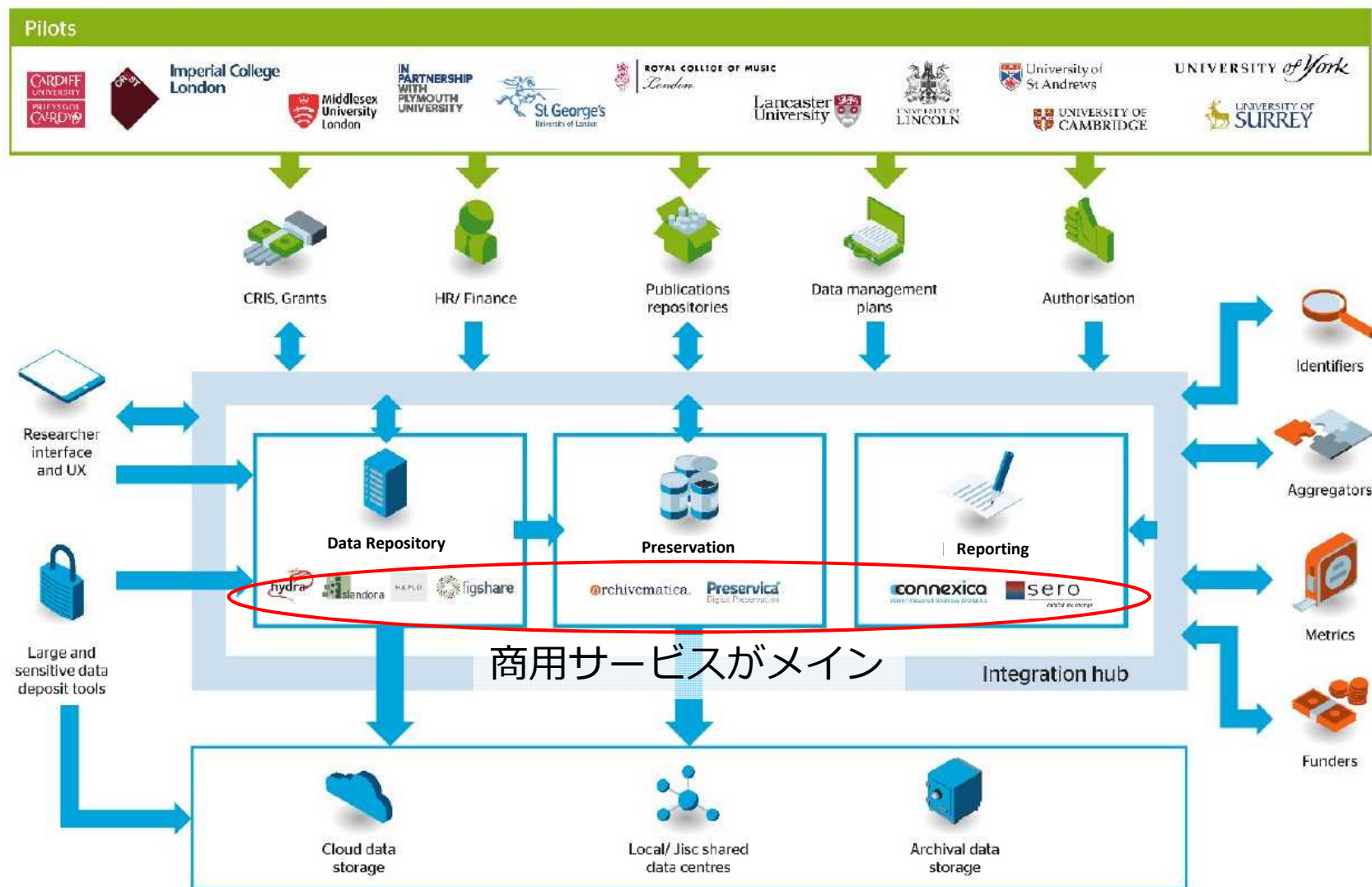
④

基盤整備の今後

イギリス：JiscプロジェクトとRDM

- **Research at risk** (2014年1月～2016年7月)
 - ポリシー、人材、基盤、インセンティブや持続的運用など、研究データ管理に関する全般的な情報をまとめたプロジェクト
 - **Research data spring** (2014年10月～2016年10月)
 - 研究データの管理や研究ワークフローをサポートするツールやサービスをコンペ形式で募ったプロジェクト
 - **Research data metrics for usage** (2015年7月～2017年10月)
 - カウンター準拠のサービスの試行運用や、その他の情報の活用などについて調査を実施したプロジェクト
- 
- **Research data shared service** (2016年1月～2018年4月)
 - 従来のプロジェクトの成果を踏襲しながら、研究データ管理に必要な基盤の整備のオプションを提示するためのプロジェクト

イギリス : Jisc RDSS Service Vision



<https://dspacecris.eurocris.org/handle/11366/611>

The CERification of the Jisc Research Data Shared Service (euroCRIS2017)より抜粋

イギリス：Jiscによるプロジェクトの進め方

- **Jiscがテーマに沿ってプロジェクトを公募**
 - 1～2年程度のプロジェクトがメイン
- **大学やベンダーが実施するプロジェクトをJiscの担当が管理**
 - 方法1：採択したプロジェクトを計画通りに最後まで推進
 - 方法2：コンペ形式で徐々にふるいにかける（最近）
- **プロジェクト成果を大学が持続的にサービス提供**
 - Jiscが直接的にサービスを運用することはあまりなかった



担当大学の努力による持続的なサービスに課題



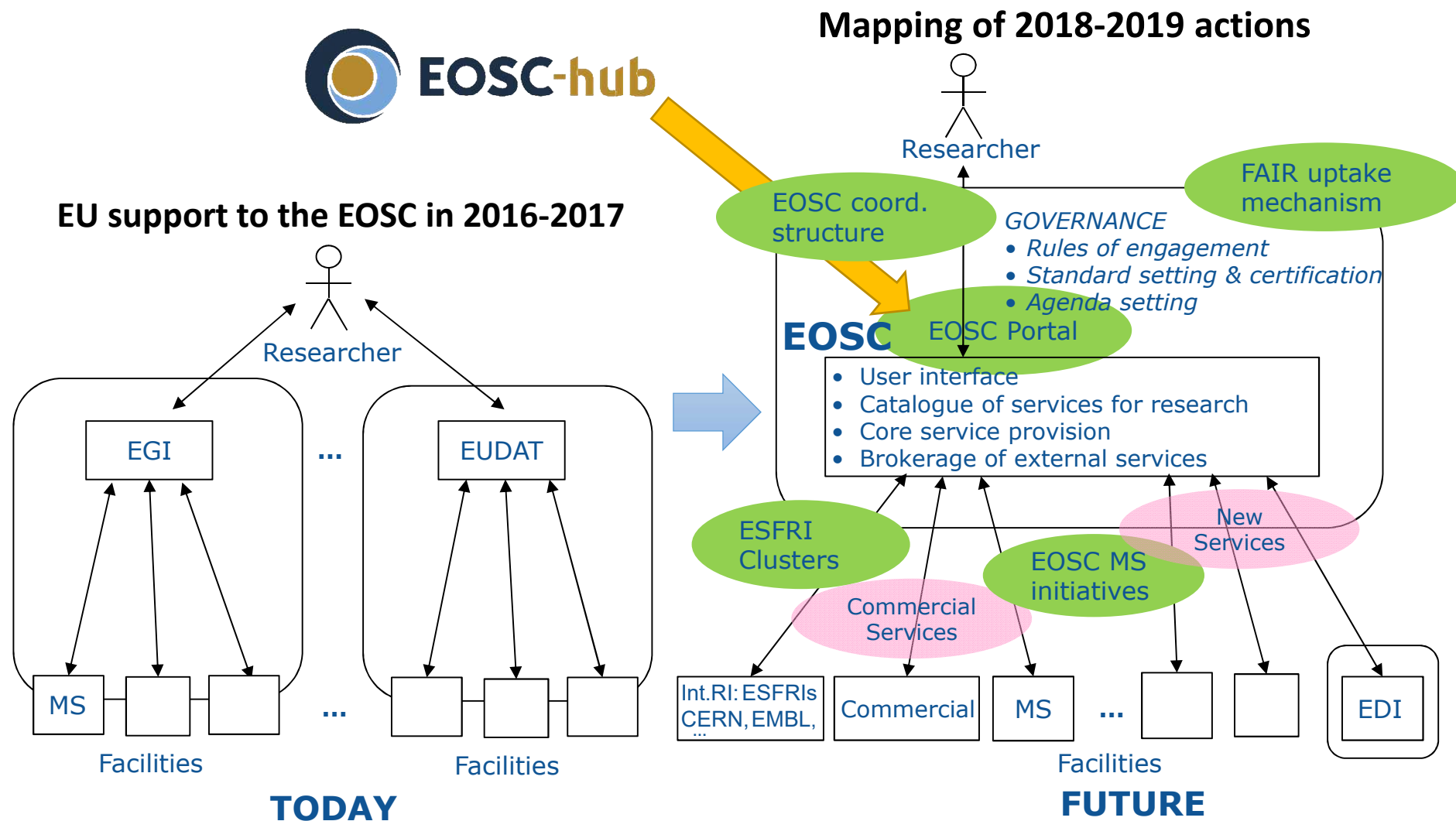
- **商用のプロダクトを活用する方向に舵切**
 - この形態については賛否両論あり

欧州：H2020におけるINFRAEOSCの公募

Horizon 2020 Work Programme 2018-2020
 part. 4. European Research Infrastructures (including e-Infrastructures)より抜粋

Research Infrastructures	2018	2019	2020	Total	Single grant	Deadlines
CALL H2020-INFRAEOSC-2018-2020	142	45.5	84	271.5		
INFRAEOSC-01: Access to commercial services	12			12	from 11 to 12 M€	22 Mar 2018
INFRAEOSC-02: Prototyping new services		28.5		28.5	from 5 to 6 M€	29 Jan 2019
INFRAEOSC-03: Integration & consolidation			79	79		mid 2020
INFRAEOSC-04: Connecting ESFRI RI through clusters	95			95	from 6 to 24 M€	22 Mar 2018
INFRAEOSC-05: EOSC governance	(a)	10		10	up to 10 M€	19 Apr 2018
	(b)	15	15	30	from 5 to 6 M€	21 Nov 2018
	(c)	10		10	up to 10 M€	19 Apr 2018
INFRAEOSC-06: portal & thematic clouds	(a)		2	2	up to 2 M€	20 Mar 2019
	(b)			5		mid 2020

欧州 : European Open Science Cloud



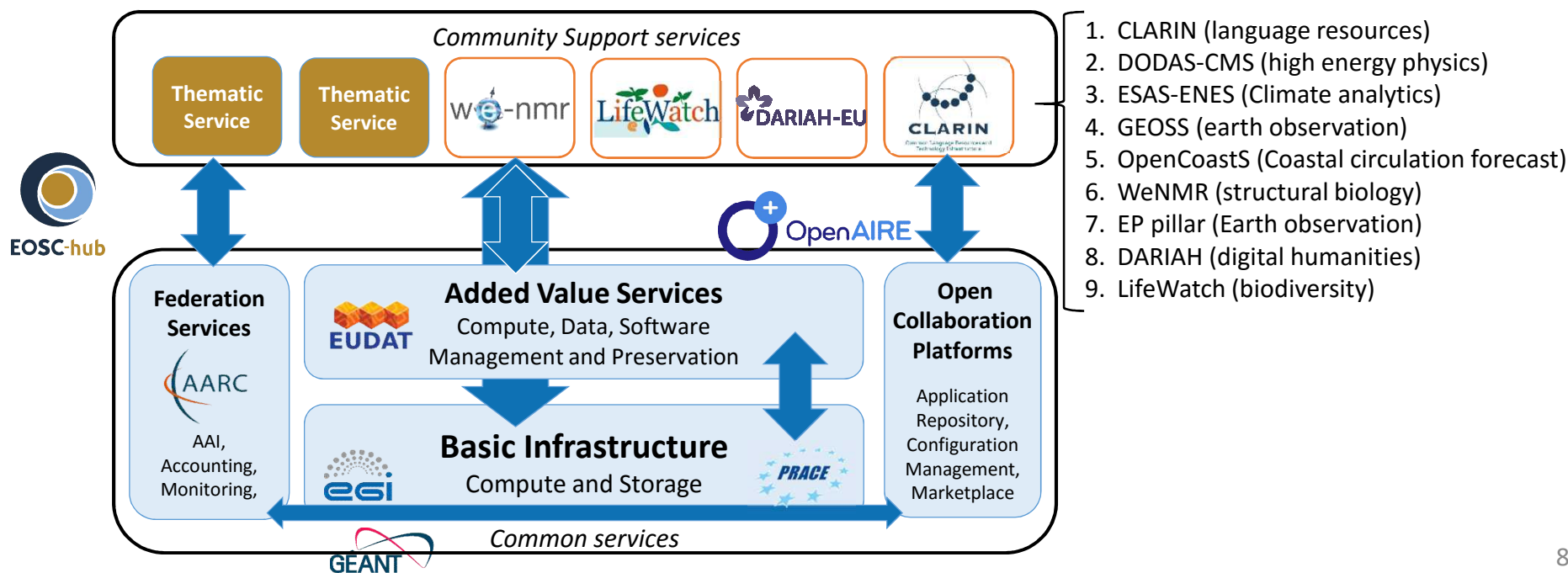
<https://eoscpilot.eu/eosc-service-architecture-and-portfolio-13-15-sept-pisa-agenda>
Current and future EU support to the EOSC and the Commission expectations after the EOSC Summitより抜粋

欧州：EOSCの状況

- 連携を期待されているサービス運用者からは、EOSCの実態がよくわからないという意見もあった



- **予算枠が具体的に確定し方向性が徐々に明確に**
 - 従来のサービスアーキテクチャー + EOSC-hub
 - EOSC-hub (€30M/3年) とOpenAIRE Advance (€10M/3年) から明示
 - より具体的なサービス連携は長期戦？



オーストラリア：研究インフラ

国家監査委員の指摘を受けて組織された調査委員会の報告



The Principles

- iii. Continuing and predictable funding for programs by shifting from the ad hoc to the sustained.
- iv. A coordinated whole-of-government approach by shifting from an unsystematic funding process.
- v. Whenever funding is provided for research, set aside appropriate additional funding for investment in infrastructure to support that research.**
- vi. Focus on Australia's research strengths and on agreed National Science and Research Priorities.
- vii. Focus on collaboration: NCRIS has clearly demonstrated the benefits of a national networked and collaborative approach.
- viii. Focus on eResearch infrastructure as the foundation for research in all disciplines.**

- ANRIF (The Australian National Research Infrastructure Fund)と呼ぶ組織を作って運用
 - どのくらいの予算を、どのような方法で確保すべき
- という詳細な内容も含め、研究力を高めるうえでの研究インフラへの効果的な投資や、Eインフラへの投資の重要性についても強く言及。

オーストラリア：組織連携の強化



の連携 (2017/18 Business Plans)

連携の狙い

1. A world leading data advantage
2. Innovation is accelerated
3. Collaboration for borderless research
4. Enhanced translation of research.

具体的な連携プログラム

- *Research Domain Program*
研究分野や研究コミュニティに必要とされる研究データインフラの提供
- *Research Data Platforms*
研究データや情報科学に必要な計算機資源、ストレージ、サービス基盤のサポート
- *Sector-wide Support and Engagement*
データの蓄積や共有に関するポリシーの策定、国際連携、人材育成の強化

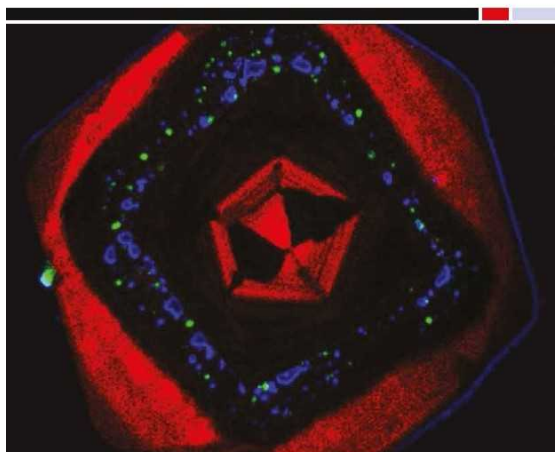


Australian Research Data Cloud (ARDC)

オーストラリア : ADRC

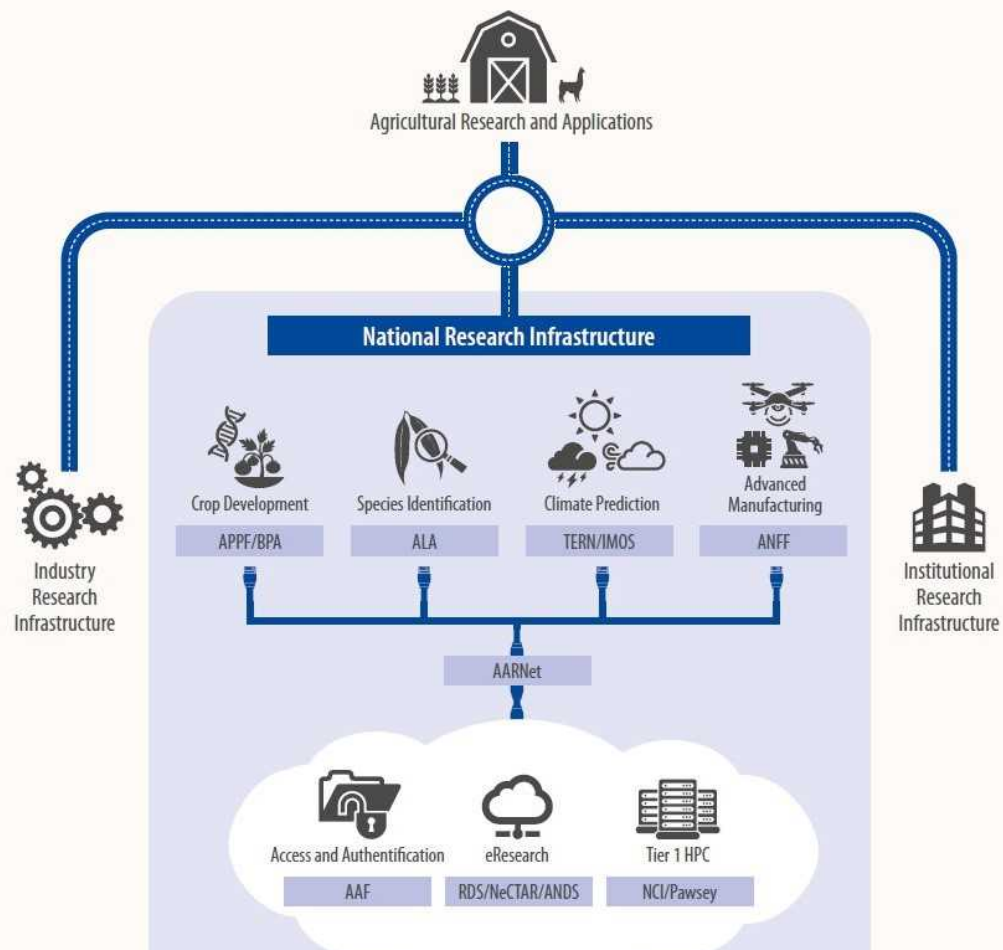


2016 NATIONAL RESEARCH
INFRASTRUCTURE
ROADMAP



ARDCについて言及

Figure 3: National research infrastructure being used by a research domain, in this case to increase agricultural productivity and sustainability



サービス連携の構成はEOSCとほぼ同じ

アメリカ

- **National Data Service** 

- NDS Labs Workbench
 - 研究データ解析のためのクラウド環境
- NDS Share
 - ラストリゾートとしてのデータリポジトリ

- **Center for Open Science** 

- Open Science Framework (OSF)
 - 研究プロジェクトでのデータ共有と公開システム
- プレプリントリポジトリ
 - 法学、図書館情報学、地球科学、栄養学、古生物学、スポーツ科学、社会科学、心理学、農学、工学、インドネシアなど
- トレーニングコース
 - OSFだけではなく、オープンサイエンスや研究データ管理に関する普及啓もう



- それぞれのサービス自体は魅力的なものが多い
- ナショナルレベルや組織横断的な活動は少ない

国際的な動向から見えてくるもの

• サービスレイヤ

このスタックが
デファクト

共通ディスクバリサービス
分野別サービス
共通サービス
クラウド・HPC
VOプラットフォーム
認証フェデレーション
ネットワーク

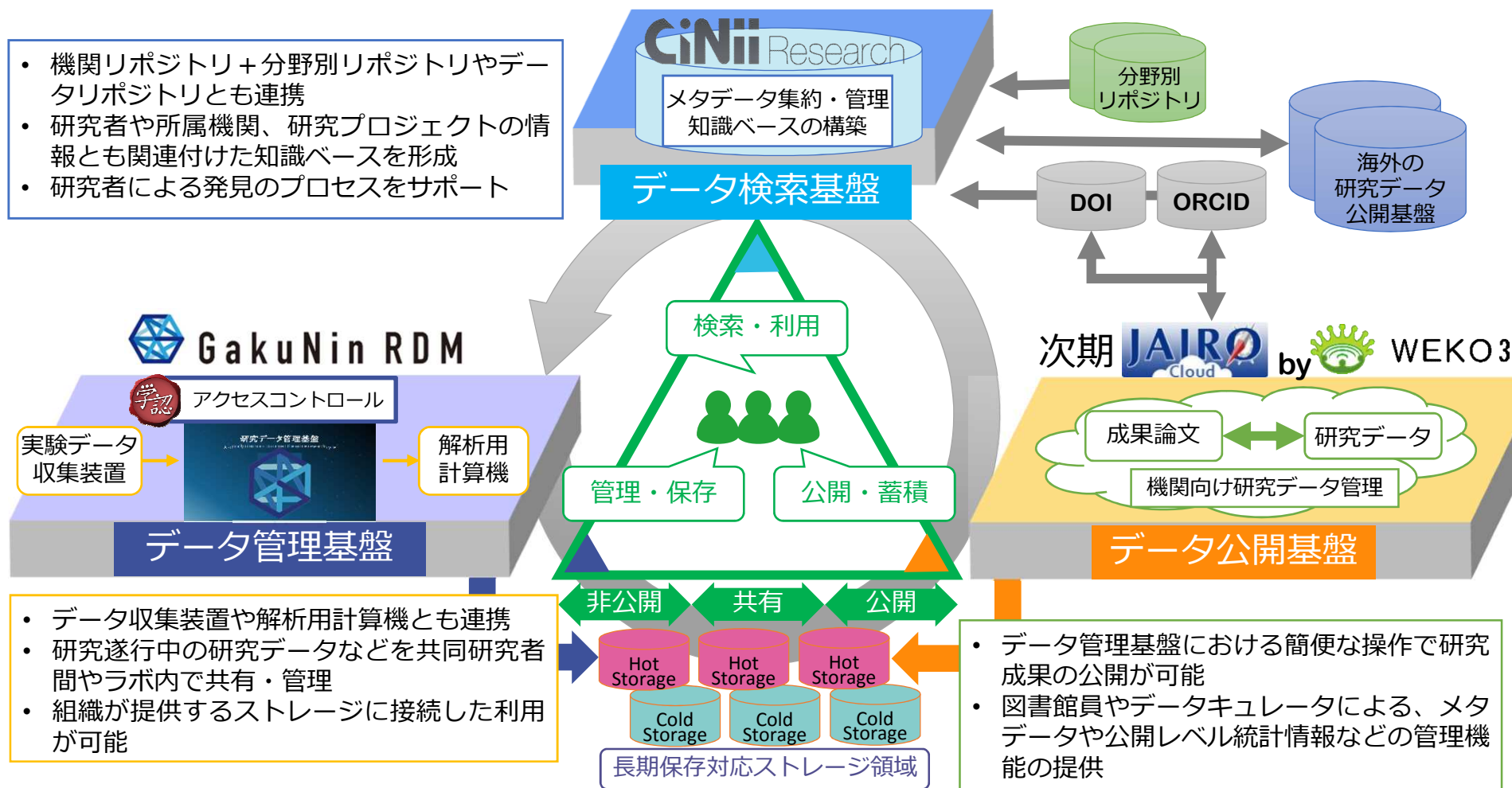
• 基盤整備のための連携

- サービス間連携：サービスレイヤ間、分野間
- 組織間連携：予算措置、共同ビジネスプラン

既存のEインフラを繋げて利便性や費用対効果を向上

これを政府レベルで積極的に推進

オープンサイエンス対応 – 研究データ基盤



研究データ基盤開発の課題

• オープンアクセス

- 論文リポジトリ、論文検索
 - 既知のフォーマットと想定可能な利用方法
⇒ 便利機能を追加

• オープンサイエンス

- 研究データを対象としたサービス
 - Variety : 研究データそのものも扱い方も多様である
 - Critical : 研究データは日々扱われる (特に管理基盤)

柔軟性・拡張性

仕様が決まらない

可用性・信頼性・保守性

止まるとすごく怒られる



システム開発・運用として

- アプリケーション内部のマイクロサービス化
- CI/CDによるDev/Opsの実現

(従来の事業形態とは全く異なる)

公開基盤のシステム構成と外部連携

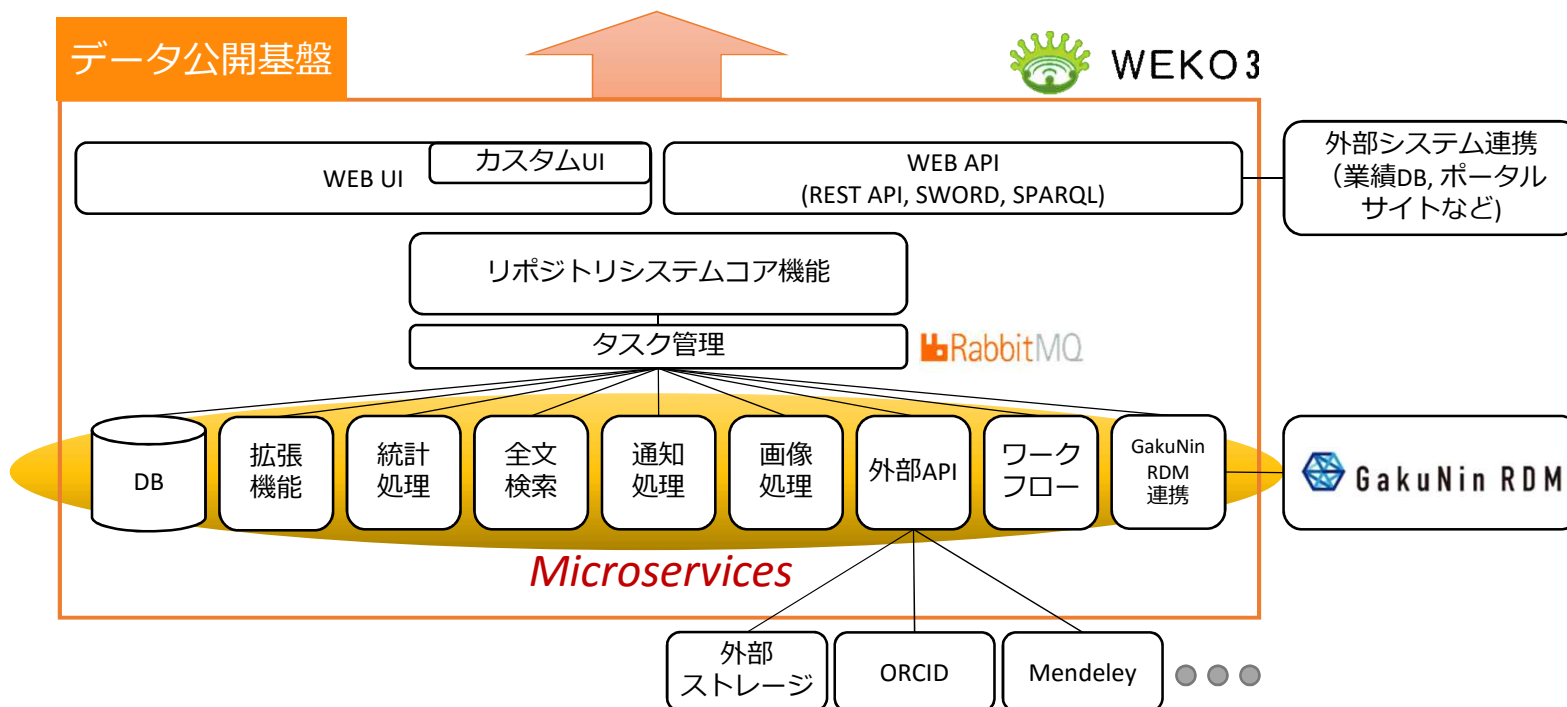
機関リポジトリ

- 論文、紀要
- デジタルアーカイブ
- 研究データ
- 教育コンテンツ
- Etc.

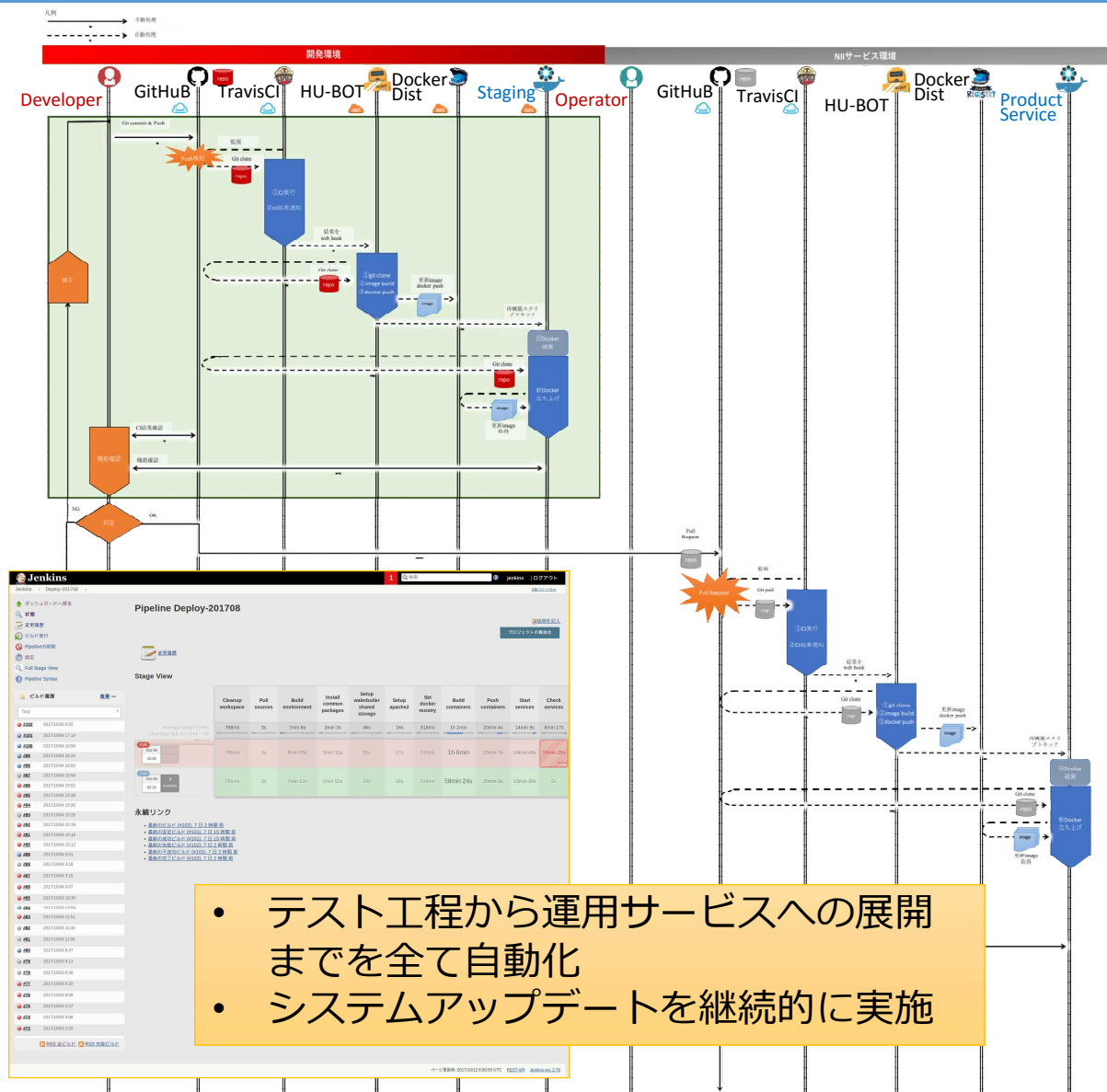
分野別リポジトリ

- 論文、テクニカルレポート
- 研究データ、サプリメンタルデータ

付加価値サービス
(統計情報, 学内システム連携など)

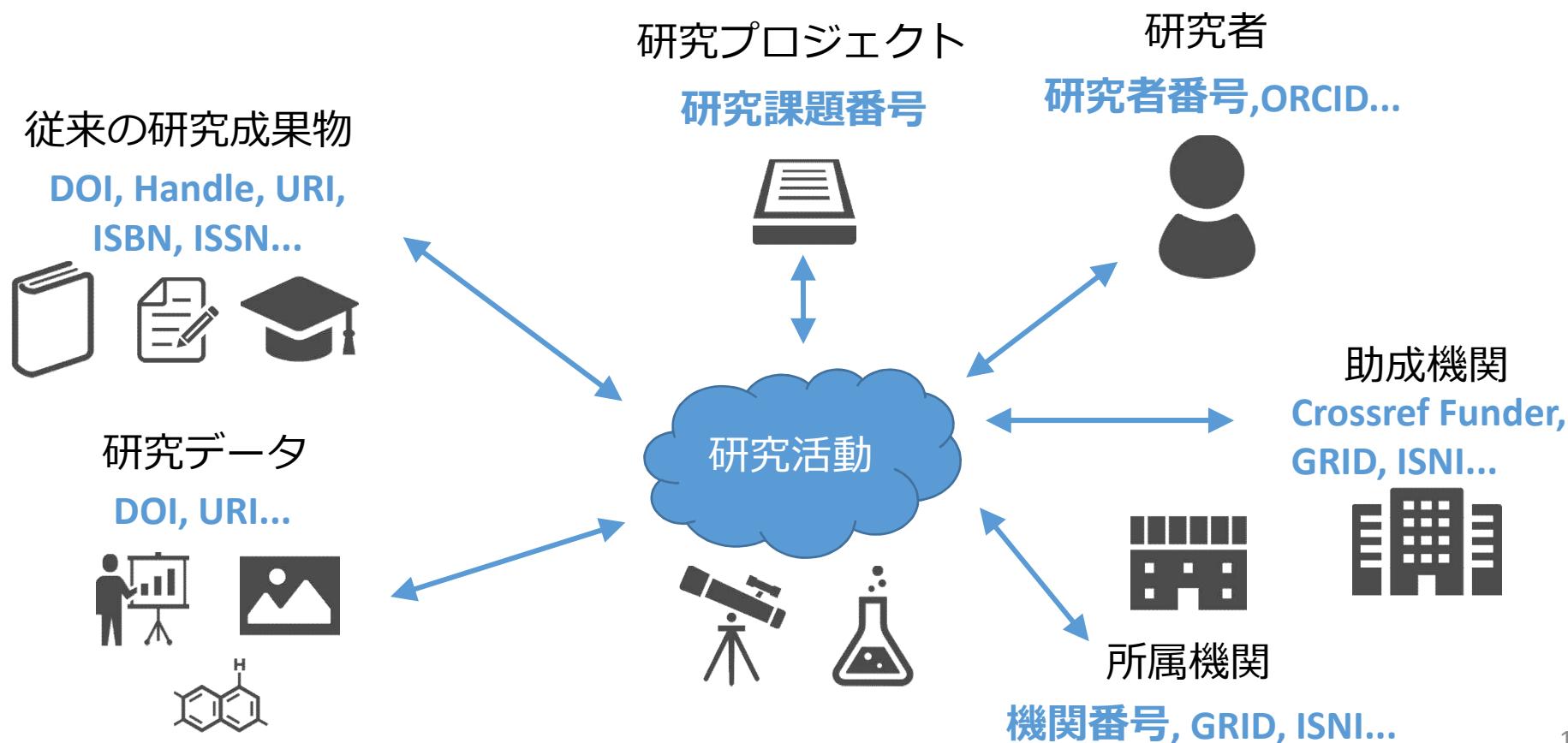


管理基盤の開発・運用フロー

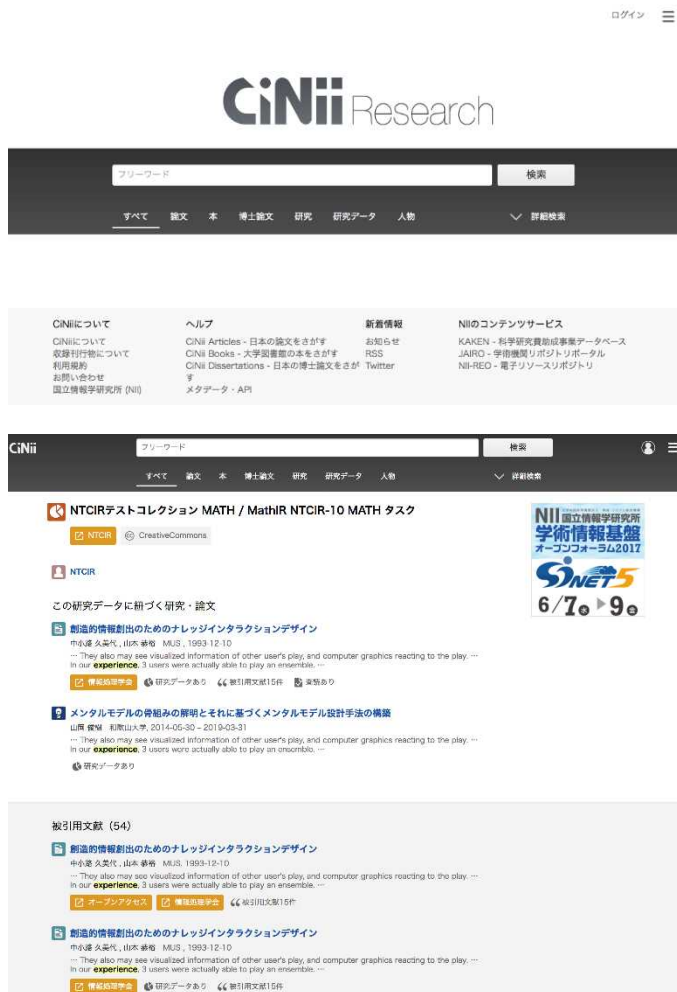


検索基盤のミッション

- 公開基盤や外部DB上の研究データのメタデータを対象とした検索
 - タイトル・作成者・作成日時・フォーマット等
- 研究活動を取り巻く情報のつながりを利用した発見
 - 文献・データ・研究者・研究課題・助成機関等



CiNii Research

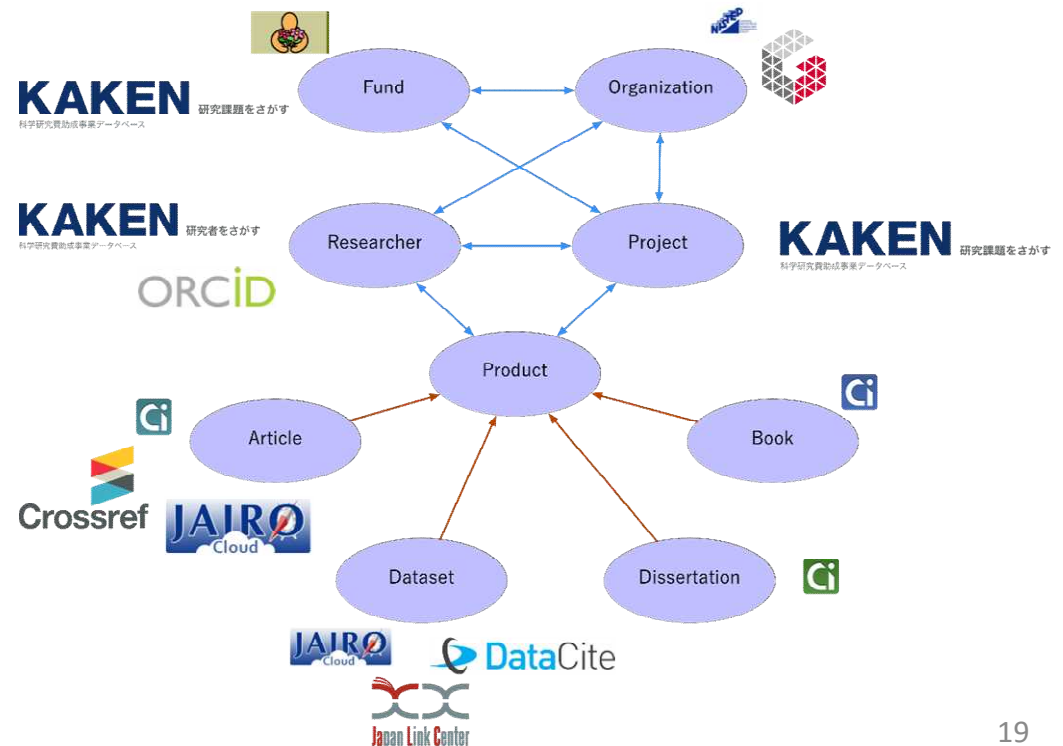


• NIIナレッジグラフの構築

- 多様なデータベースからの情報集約
- エンティティ・リンクへの再編

• ユーザーインターフェースの提供

- 文献や研究課題を通じた研究データの発見を支援



今後の展開

2017年開発：対象データの収集とAPI設計、UIの初期実装
 2018年開発：ナレッジグラフの構築とUIの改善、運用基盤構築

・国内リソースへの展開

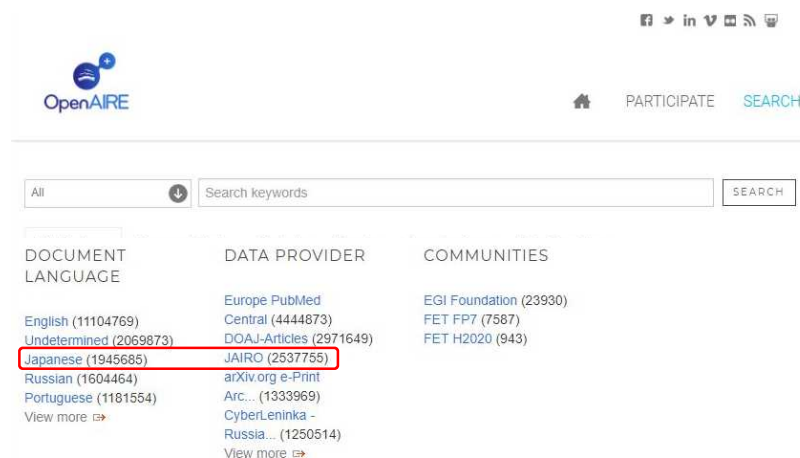
- ・ 2019 試行版提供
- ・ 2020 正式版公開

・多様なドメインとの連携

- ・ 国内研究分野DB
- ・ サービスレベルの連携
 - ・ API提供・利用

・国際連携

- ・ OpenAIRE
- ・ データレベルの連携
 - ・ ナレッジグラフの相互利用



公開基盤のミッション

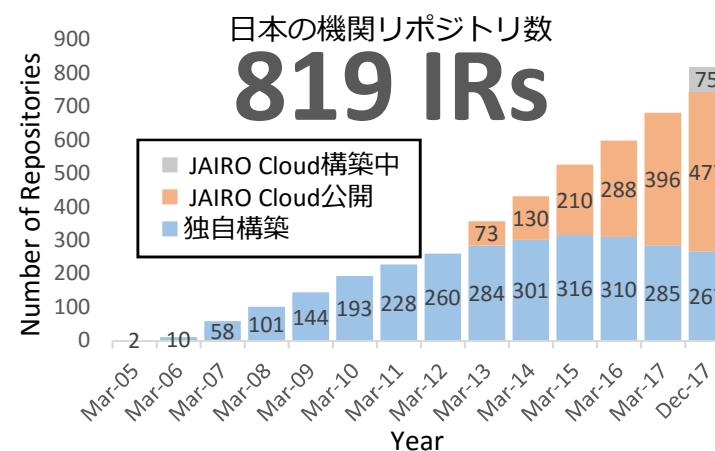
・文献リポジトリとして

- ・ 機関や研究費助成機関からの義務化に対応
 - ・ 機関内における異なるワークフローに対応可能
 - ・ 登録や公開のワークロードを軽減できる機能提供
- ・ JAIRO Cloudサービスとしての成熟
 - ・ 国内の全機関が利用してもサービスの可用性・信頼性・保守性を担保

・データリポジトリとして

- ・ 論文と関連したエビデンスデータの簡便な登録
- ・ 多様な利用形態や機能要件に対応
- ・ スケーラビリティを確保

どのような方法で
実現しているか？





WEKO3

• 現状のWEKO2

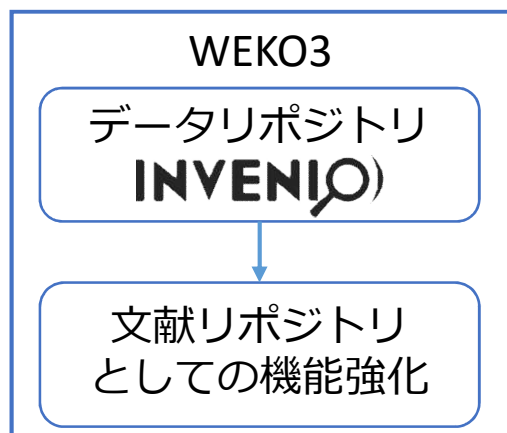
- 文献の扱いを中心とした機能を提供
- 実ユーザ（JAIRO Cloud）の声を常に反映し高機能化



研究データ対応

• 新しいWEKO3

- データリポジトリとして実績をもつInvenio3をベース
- WEKO2までの機能群をInvenio3に踏襲



Invenio3の洗練されたアーキテクチャー上で
公開基盤に必要とされる機能群を実現
(Invenio3と管理基盤は同様のシステム構成)

- ↳ 運用やシステム開発の効率化
- ↳ 拡張性を生かしたドメイン展開

今後の展開

2017年開発：WEKO2機能、基本UIの開発

2018年開発：ワークフロー機能の強化とUIの改善、エビデンス
データ登録機能の開発、運用基盤環境の構築

• JAIRO Cloudへの展開

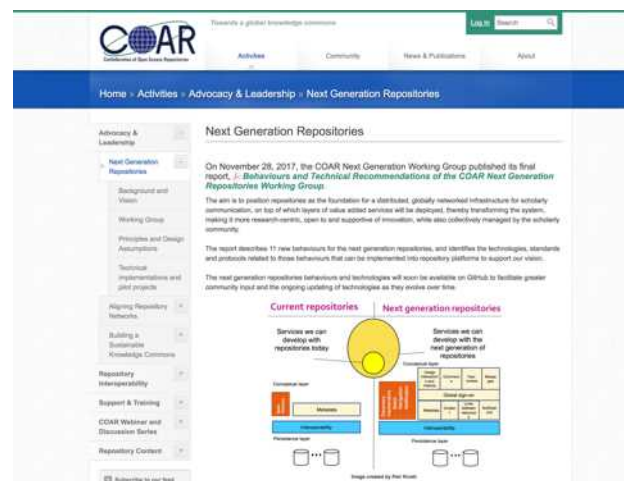
- 2019 試行運用
- 2020 正式運用

• 多様なドメインとの連携

- 分野別リポジトリ
 - 分野特有の機能実装
- 付加価値サービス
 - API連携・利用

• 国際連携

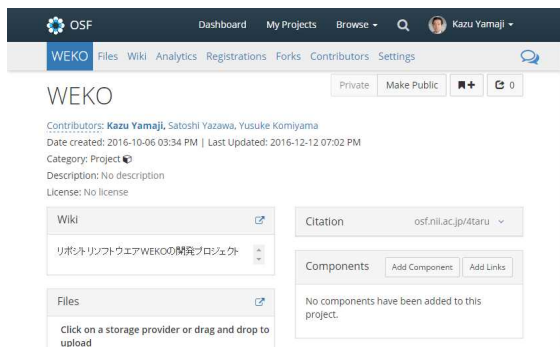
- COAR（特にNGR機能の実装）
- CERN & Invenio Community



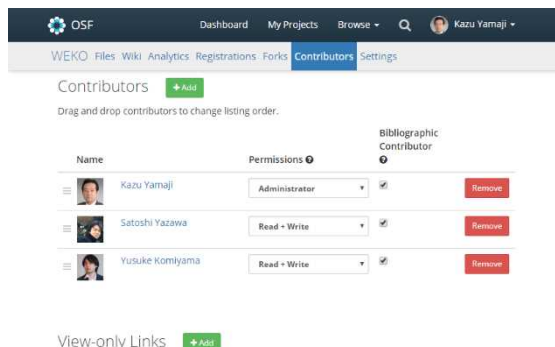
GakuNin RDM

米国COSが開発したOpen Science Frameworkを拡張

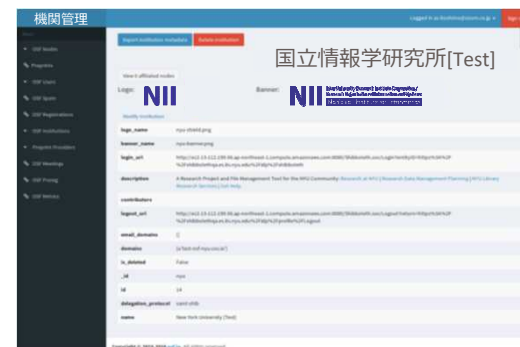
研究プロジェクト単位で
ファイルなどを管理



学認と連携しVO(仮想組織)
メンバーでファイルを共有



機関提供のストレージを利用し
研究証跡を保存・保護



フロントエンドはNIIが提供

研究データ管理基盤

機関毎に準備

クラウドストレージ

NIIストレージ

大学・研究機関毎の既存のクラウドストレージの
事情に合わせてプラグインをカスタマイズ

NII提供の最小限の
デフォルト領域



パブリッククラウド
(プロバイダーDC)



プライベートクラウド
(オンプレミス環境)

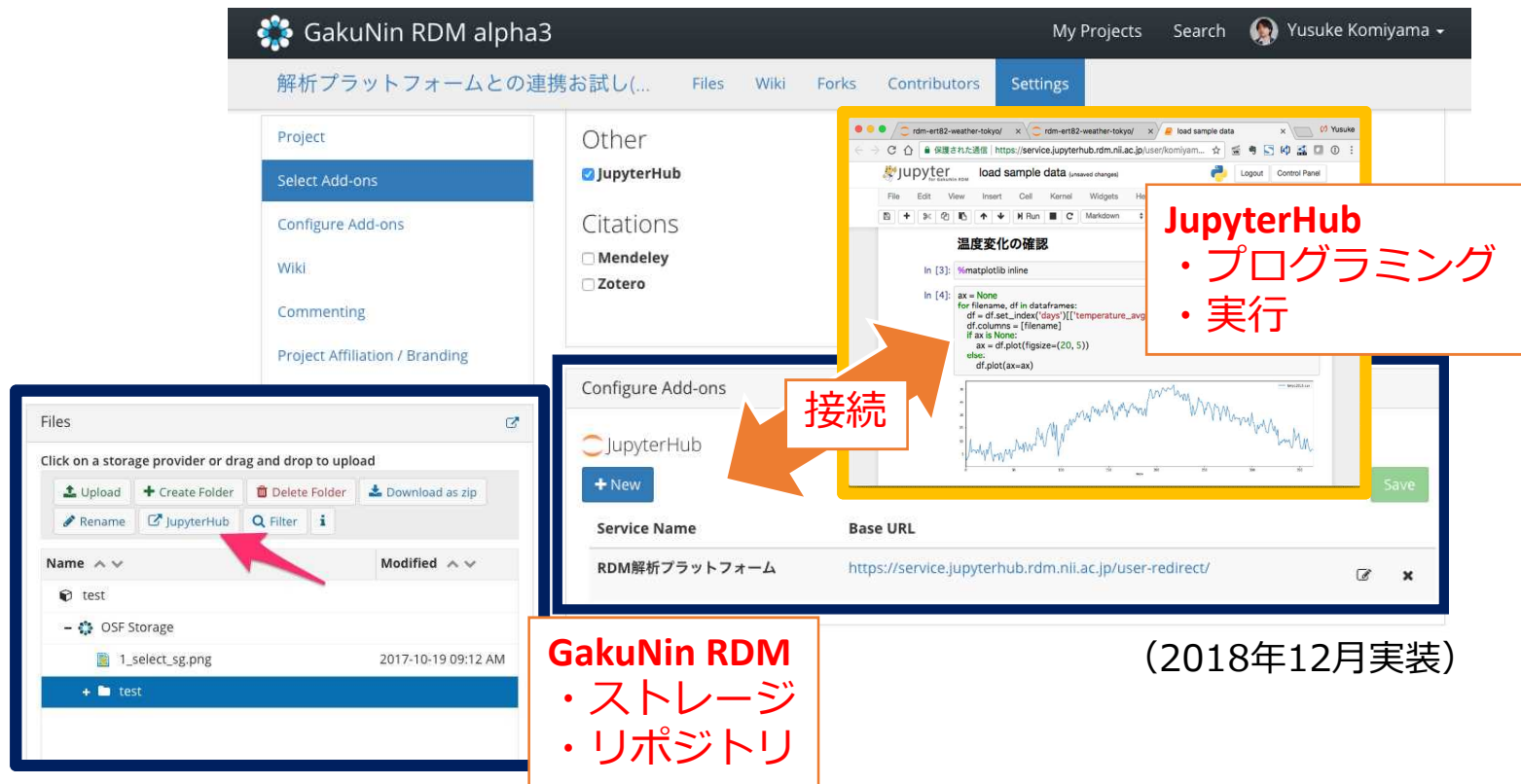


パブリッククラウド
(プロバイダーDC)

開発機能の紹介

- 新規アドオン
 - 新規外部ストレージ対応
 - ownCloud, S3 Compatible Storage, OpenStack Swift
 - データ解析ソフトウェア連携
 - JupyterHub連携
 - プラグインSDK開発
- 研究データ管理
 - 研究証跡保存
 - メタデータ管理
 - ワークフロー対応
- 機関向け管理
 - プラグイン選択
 - 統計
 - 機関テンプレート

データ解析ソフトウェア連携



JupyterHub

- ・プログラミング
- ・実行

接続

GakuNin RDM

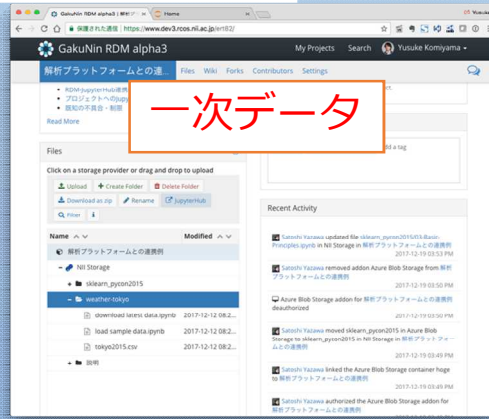
- ・ストレージ
- ・リポジトリ

(2018年12月実装)

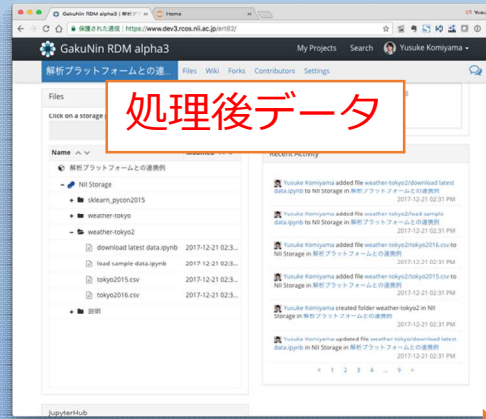
- データ解析プラットフォームJupyterHub用のGakuNin RDMアドオンを開発
- 連携によりシステム間でのデータ入出力が容易に
- 双方のアプリケーションでの認証は学認に対応

データ解析ソフトウェア連携の利用イメージ

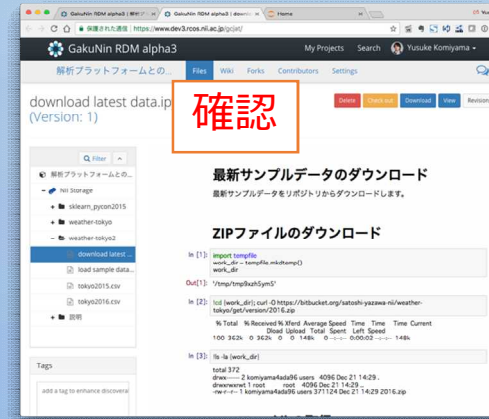
GakuNin RDM側



(1) ディレクトリをGakuNin RDMから JupyterHubへ出力

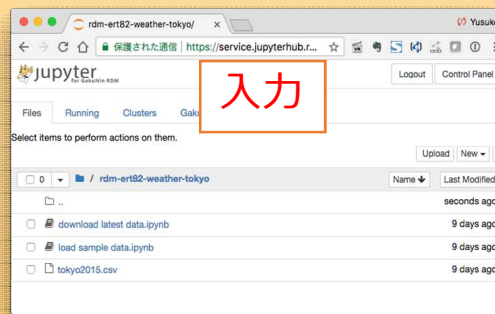


(5) GakuNin RDMでファイルを確認

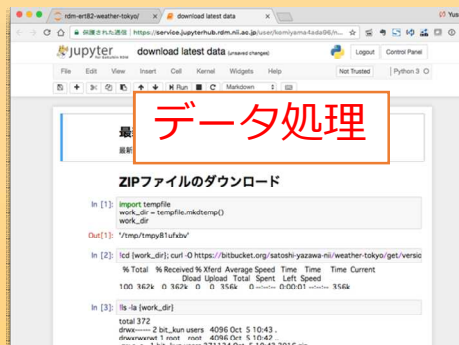


(6) IPython Note Book (ipy nb) 形式はプレビューにも対応

JupyterHub側



(2) JupyterHubでファイルを確認

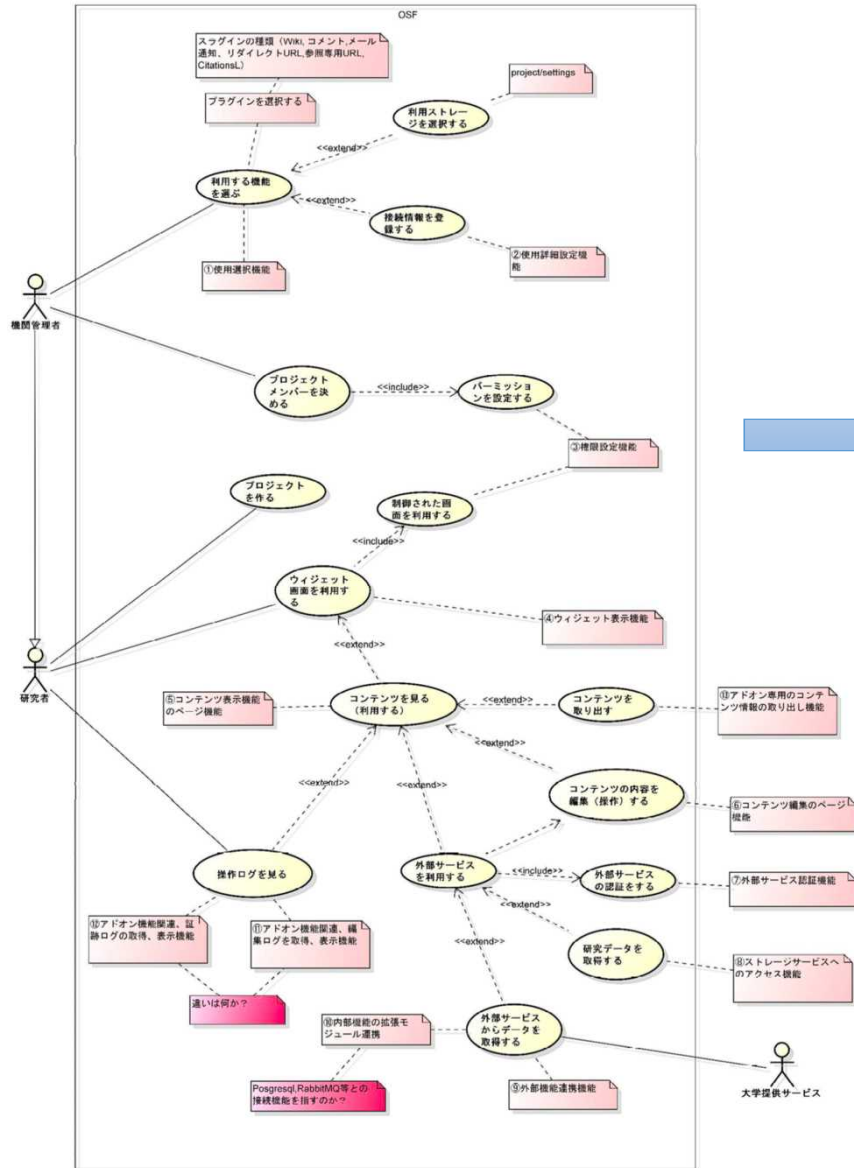


(3) プログラミング & 実行



(4) 実行結果をJupyterHubからGakuNin RDMへ出力

プラグインソフトウェア開発キットの提供



用途

- 開発者向けにユースケースを想定
- GakuNin RDMプラグインSDKを開発



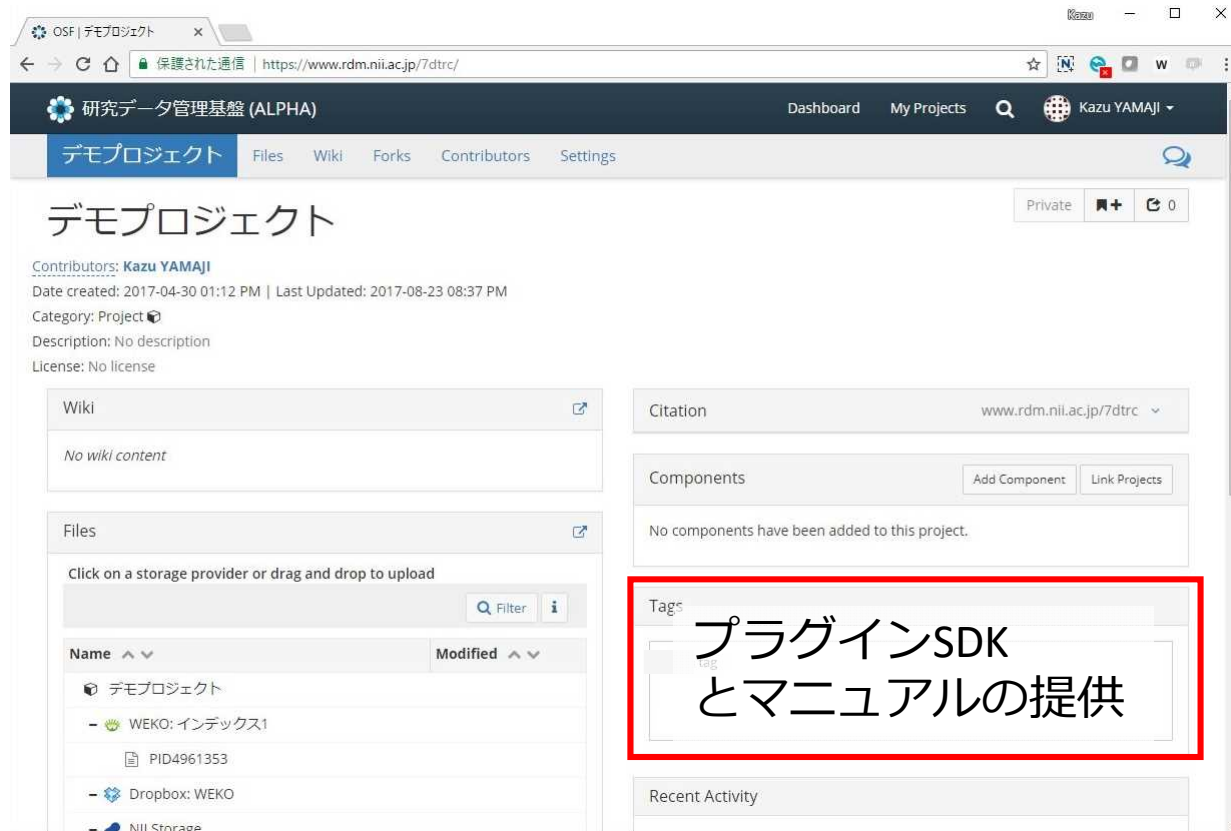
各機関や各研究分野のサービスと接続するために、Pythonのソースコードをテンプレートとして出力



ユースケースの例

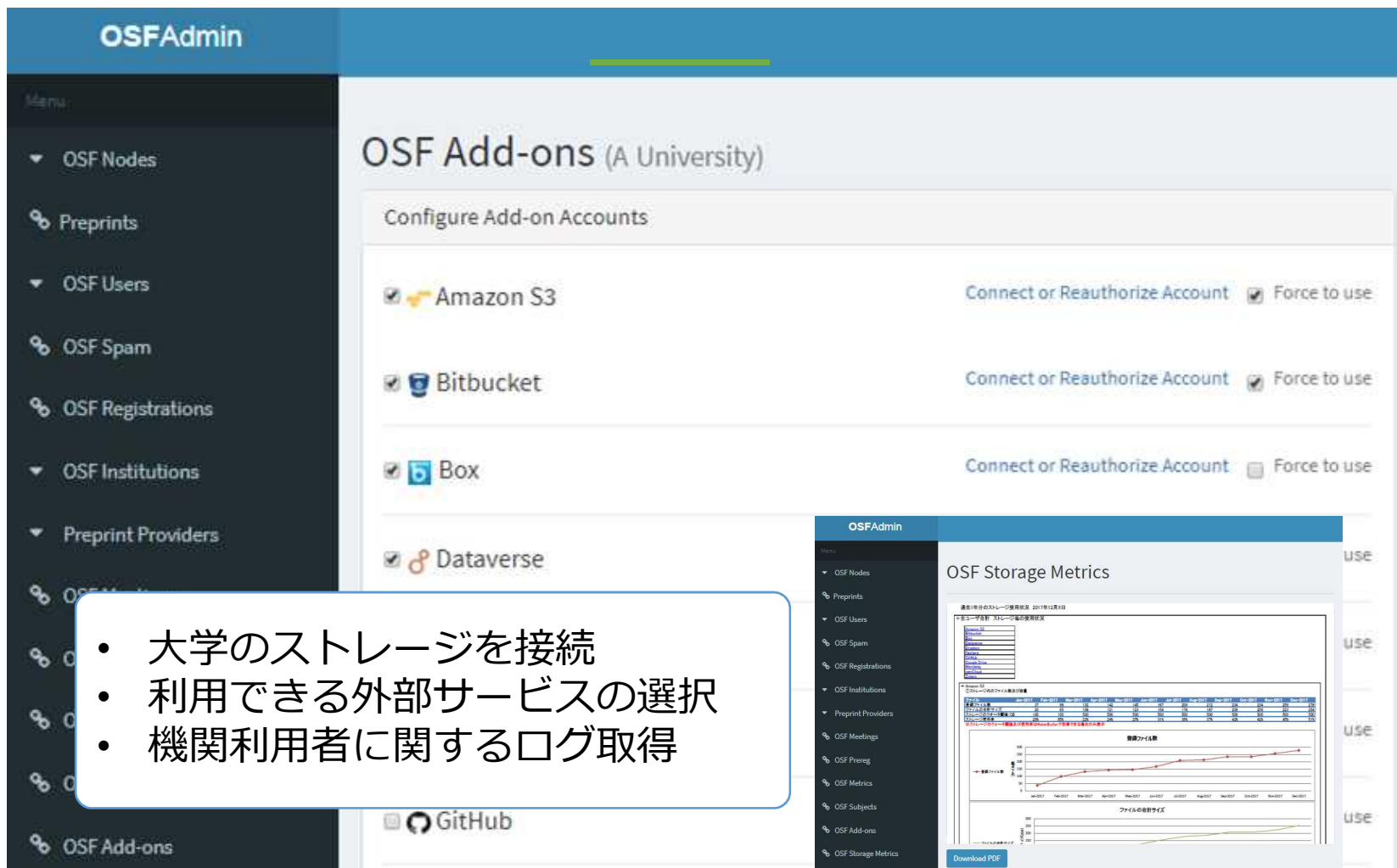
- 利用する機能を選択
- 利用ストレージを選択
- 接続情報を登録
- プロジェクトメンバーの決定
- パーMISSIONの設定
- プロジェクト作成
- ウィジット画面利用
- 制御された画面を利用
- コンテンツ閲覧・利用
- コンテンツ出力
- コンテンツ編集・操作
- 操作ログの閲覧
- 外部サービスの利用
- 外部サービスの認証
- 外部サービスからのデータ取得

プラグインSDKを埋め込んだイメージ



- 大学・研究機関と連携し目的特化型の独自プラグインの開発
- 研究データ基盤の開発者コミュニティの形成

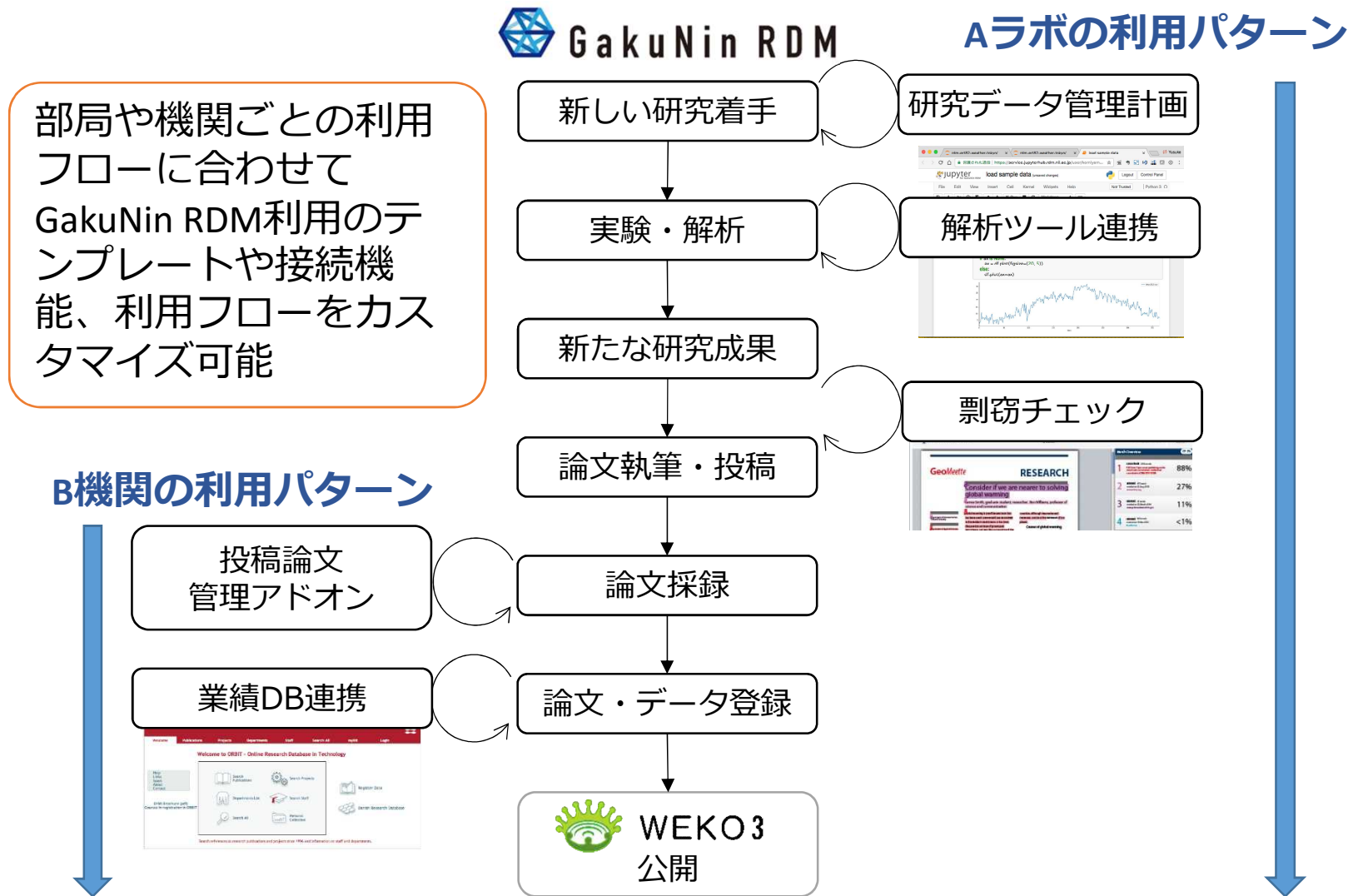
機関向け管理機能



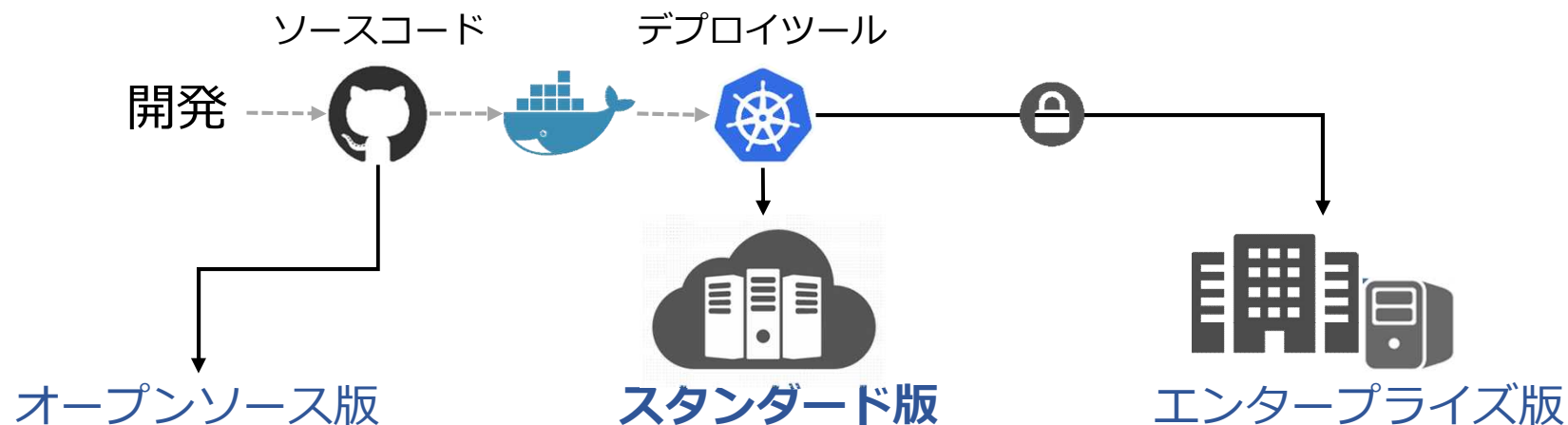
The screenshot displays the OSFAdmin interface. The top navigation bar is blue with the 'OSFAdmin' logo. A dark sidebar on the left contains a 'Menu' with various options: OSF Nodes, Preprints, OSF Users, OSF Spam, OSF Registrations, OSF Institutions, Preprint Providers, OSF Meetings, OSF Prereg, OSF Metrics, OSF Subjects, OSF Add-ons, and OSF Storage Metrics. The main content area is titled 'OSF Add-ons (A University)' and features a 'Configure Add-on Accounts' section. This section lists four add-ons: Amazon S3, Bitbucket, Box, and Dataverse. Each add-on has a checkbox (all are checked), a 'Connect or Reauthorize Account' button, and a 'Force to use' checkbox. Below this, the 'OSF Storage Metrics' dashboard is visible, showing a table of storage usage data and two line graphs: '蓄積ファイル数' (Accumulated File Count) and 'ファイルの総サイズ' (Total File Size). A 'Download PDF' button is located at the bottom of the metrics section.

- 大学のストレージを接続
- 利用できる外部サービスの選択
- 機関利用者に関するログ取得

機関の利用モデルに沿った展開



GakuNin RDMの提供方法



- 機関が独自にGakuNin RDMを運用したい場合
- ソースコードとドキュメントを提供
- 運用のノウハウを共有

- パブリッククラウド上のSaaSとしてNIIが提供
- 機関が用意するストレージを接続して利用可能
- 学認による認証

- 機関のポリシー上スタンダード版ができない場合
- デプロイコードの提供、あるいはNIIが直接デプロイ
- 機関のIdPによる認証

機関のポリシー等の制約などにも対応できるようにいくつかの運用オプションを検討中

大学・研究機関との実験計画

- クローズドテスト（第1回） 2017年3月実施完了
目的 : 大規模大学のIT基盤管理部門担当者を中心に大学/機関へのシステム導入時に問題になることが予想される内容を中心にテストを実施
- クローズドテスト（第2回） 2017年10月実施完了
目的 : 中小大学、機関のIT基盤管理部門での導入に関連する知見の収集、及びラボレベルでの研究者、データエンジニア等に利用してもらい、様々な分野の研究フローにおける要望や、問題点に関する知見を得る。
- **クローズドテスト（第3回） 2018年3月実施予定**
目的 : 2017年度開発機能を含め、大学/機関の様々な職責の人による、評価と問題点などの知見を得る。
- 機関内パッケージ試用テスト 2018年4月以降
目的 : 機関内限定利用、あるいは内部での試験。
特定分野向け、あるいは特定機関用アドオン開発環境構築、機能開発用
- 限定オープン利用テスト 2018年4月以降
目的 : 試用対象者向けに常時利用可能な環境を公開し、実運用に近い形での利用をお願いし、その中での要望、不備などの洗い出し。

実験からの発展

• 小規模ラボでの利用

- 医学、地球科学などのラボレベルでの利用実証
- 社会調査データにおける具体的な管理やグループ共有

• 基盤センターレベルでの採用

- いくつかの大学の基盤センターと、具体的に機関としての利用のための調整開始

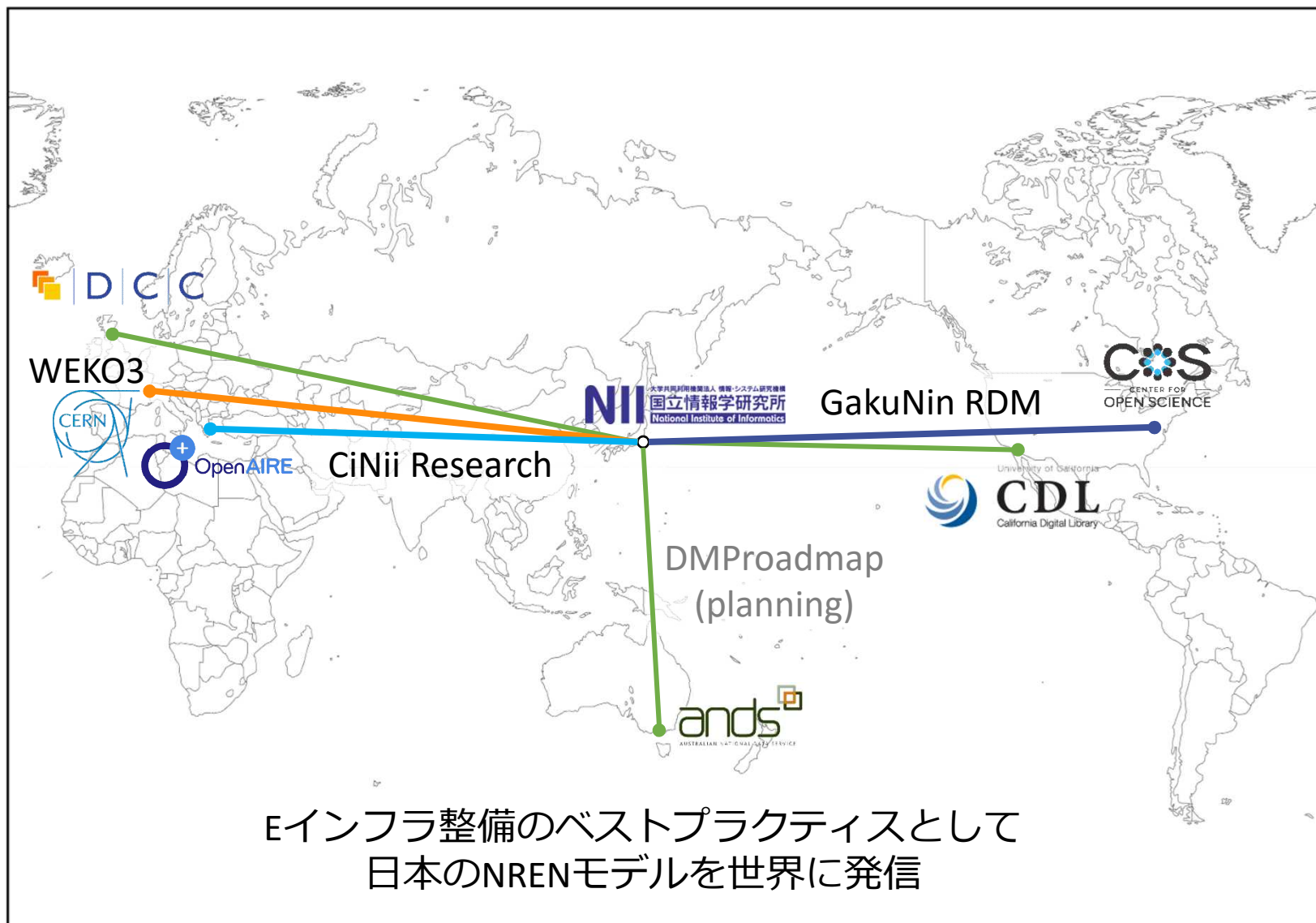
• 学内ポリシーとのすり合わせ

- 学内における研究データ管理ポリシーの策定と合わせながら、サービスの機能要件を検討

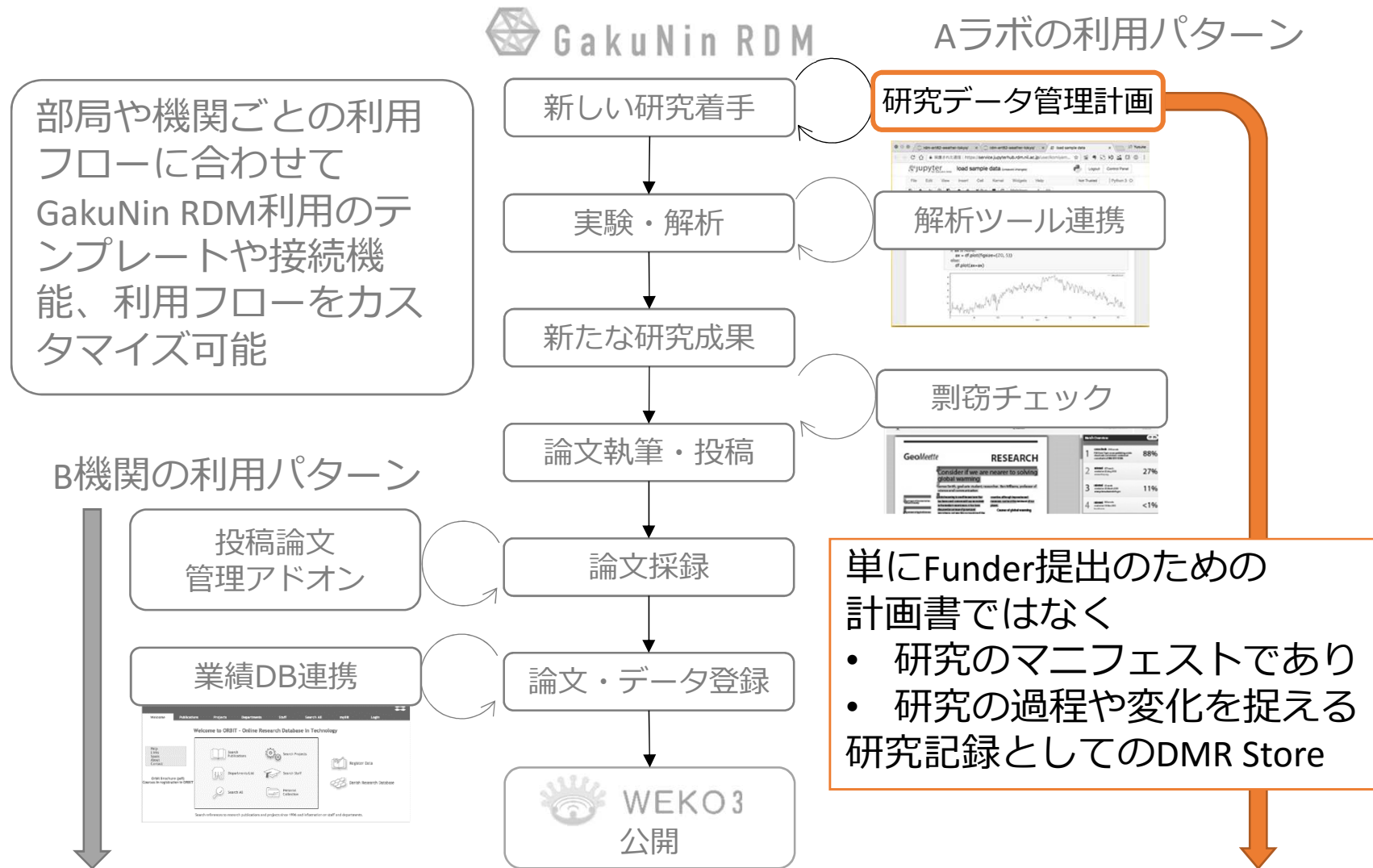
• 研究不正対策としての機能実証

- 論文とエビデンスデータを管理する機能に特化した利用をワークフローとともに実証

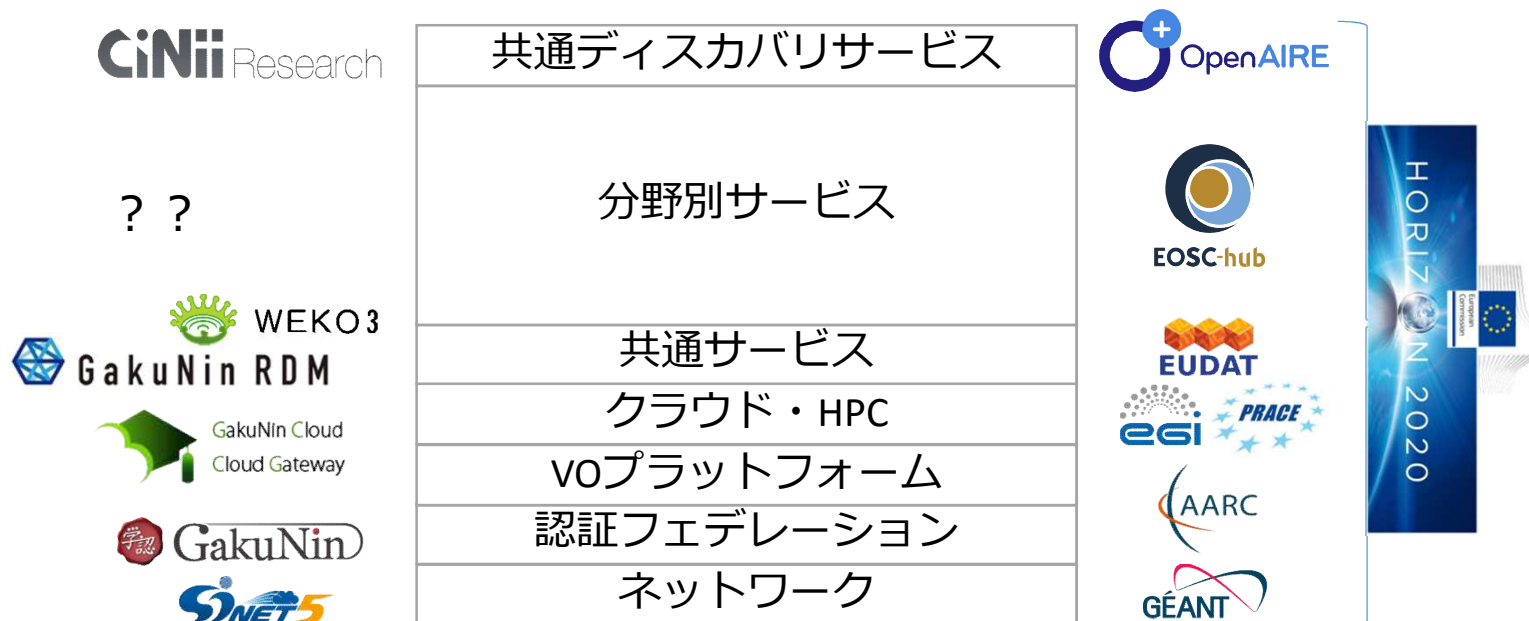
研究データ基盤開発の国際連携



Data Management Records Store



日本と他国の比較から見える課題



- 分散する分野別プラットフォームをどう繋いでいくか？
- 他の研究インフラとどう積極的に繋げていくか？
- ポリシーメーカーとインフラ利用をどう繋げていくか？
- 日本としてEインフラの強化をどう位置付けるか？

RCOS
yamaji@nii.ac.jp